






ARTICLE



<https://doi.org/10.1038/s41467-020-19170-5>

OPEN

The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s

Hale Tunbak^{1,7}, Rocio Enriquez-Gasca ^{1,7}, Christopher H. C. Tie², Poppy A. Gould¹, Petra Mlcochova³, Ravindra K. Gupta³, Liane Fernandes¹, James Holt¹, Annemarte G. van der Veen^{4,6}, Evangelos Giampazolias⁴, Kathleen H. Burns ⁵, Pierre V. Maillard ¹ & Helen M. Rowe ¹ 

The Human Silencing Hub (HUSH) complex is necessary for epigenetic repression of LINE-1 elements. We show that HUSH-depletion in human cell lines and primary fibroblasts leads to induction of interferon-stimulated genes (ISGs) through JAK/STAT signaling. This effect is mainly attributed to MDA5 and RIG-I sensing of double-stranded RNAs (dsRNAs). This coincides with upregulation of primate-conserved LINE-1s, as well as increased expression of full-length hominid-specific LINE-1s that produce bidirectional RNAs, which may form dsRNA. Notably, LTRs nearby ISGs are derepressed likely rendering these genes more responsive to interferon. LINE-1 shRNAs can abrogate the HUSH-dependent response, while over-expression of an engineered LINE-1 construct activates interferon signaling. Finally, we show that the HUSH component, MPP8 is frequently downregulated in diverse cancers and that its depletion leads to DNA damage. These results suggest that LINE-1s may drive physiological or autoinflammatory responses through dsRNA sensing and gene-regulatory roles and are controlled by the HUSH complex.

¹Centre for Immunobiology, Blizard Institute, Queen Mary University of London, London E1 2AT, UK. ²Infection and Immunity, University College London, London WC1E 6BT, UK. ³Department of Medicine, University of Cambridge, CB2 0AF Cambridge, UK. ⁴The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ⁵Department of Pathology, John Hopkins University School of Medicine, Baltimore, MD 21205, USA. ⁶Present address: Leiden University Medical Centre, Department of Immunohematology and Blood Transfusion, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. ⁷These authors contributed equally: Hale Tunbak, Rocio Enriquez-Gasca. ✉email: h.rowe@qmul.ac.uk

Throughout evolutionary history, the human genome has been continuously bombarded with genome invasions, either from exogenous retroviruses that have infected the germline, or from new waves of endogenous transposition¹. This process has led to up to two-thirds of our genome now consisting of repetitive DNA². Transposition is important to drive genome innovation, and there is an increasing number of documented cases where transposable elements (TEs) have been repurposed, for example to act as enhancers or promoters, with functions in pathogen defense³. Regulation of the mobile genome is critical to maintain genome integrity, however, and this is largely achieved through epigenetic silencing. For example, long interspersed element-1s (LINE-1s) are targeted by histone deacetylation, H3K9me3 and DNA methylation^{4–7}. Recent data have revealed that inactivation of epigenetic silencing, particularly in human cancer cell lines leads to the reactivation of endogenous retroviruses and other TEs^{8–15}. This is accompanied by activation of cytosolic nucleic acid sensors of the innate immune system and production of type I interferons (IFN α and IFN β , reviewed in ref. 16). Such activation can overcome resistance to checkpoint blockade and promote anti-tumor immunity in mouse cancer models¹⁴, although type I IFNs can worsen cancer progression and disease outcome when they are constitutively produced¹⁷. IFN production following treatment of colorectal or ovarian cancer cells with pharmacological inhibitors of DNA methylation appears to depend on double-stranded RNA (dsRNA) sensing through the RIG-I-like receptors (RLRs), MDA5 (melanoma differentiation factor 5) and RIG-I (retinoic acid-inducible gene I) and activation of their downstream mitochondrial adaptor, MAVS (mitochondrial antiviral signaling protein)^{8,13}. Very little is understood, however, about which classes of repetitive DNA potentially produce RNA molecules that engage with the RLRs and mediate cross-talk with innate and adaptive immunity, either in physiological or disease contexts. Likewise, little is known about which epigenetic pathways are involved in transcriptional control of repetitive DNA in human adult tissues^{12,15} or in cancer cells^{10,14}. Notably, however, Alu elements that exhibit an inverted repeat configuration in the genome have been identified to form duplex RNA recognized by the dsRNA sensor MDA5^{18–20} and LINE-1 elements are implicated in IFN activation, including through cGAS (cyclic GMP-AMP synthase) and STING-(stimulator of IFN genes)-dependent DNA sensing^{21–24}.

The human silencing hub (HUSH) complex was discovered as a novel epigenetic complex responsible for position-effect variegation of integrated transgenes in human cells²⁵. It is associated with H3K9me3-dense genomic regions, including the 3 prime ends of zinc finger genes and functions to repress zinc finger genes and ribosomal DNAs^{25–27}. The HUSH complex partners with the chromatin regulator MORC2, which contains a zinc finger and an ATPase domain²⁶. Mutations in MORC2 are associated with neuropathies and these disease mutations have been found to affect MORC2 structure and HUSH function²⁸. The HUSH complex has also been discovered to mediate epigenetic silencing of murine leukemia virus through its recruitment by NP220 to the viral DNA before its integration, thereby preventing viral infection²⁹. HUSH exerts some position-dependent repression on HIV, which encodes accessory proteins that can counteract HUSH repression, through mediating degradation of HUSH complex components^{25,30,31}.

We and others have recently demonstrated the HUSH complex to be necessary to repress LINE-1 elements^{32–34}. We showed that the HUSH complex represses LINE-1s and co-regulated genes in naïve mouse embryonic stem cells³⁴ and is mainly required to regulate the youngest species-specific LINE-1s^{32,34}. Interestingly, MORC2 exerts some binding specificity for full length LINE-1s³³. Here, we reasoned that HUSH-regulated LINE-1s with full-length

mRNAs and potentially intact ORFs may have been repurposed to drive nucleic acid sensing in human adult tissues. We show here that the HUSH complex negatively regulates the type I IFN response in human cells through epigenetic regulation of LINE-1 elements. Furthermore, we provide evidence that young LINE-1s produce bidirectional transcripts that could form dsRNA.

Results

The HUSH complex regulates type I IFN signaling. We asked whether the HUSH complex regulates type I IFN induction and responsiveness in human cells by employing shRNA-mediated HUSH depletion in a HEK293 IFN reporter cell line. These cells have been described before^{35,36} and harbor an integrated lentivector expressing destabilized GFP under control of IFN-stimulated response elements (ISRE), referred to as ISRE-GFP (Fig. 1a). We verified by western blot or qRT-PCR that HUSH components were depleted by day 4 post introduction of the shRNAs (Fig. 1b) and found that the ISRE-GFP reporter becomes activated at day 6 post shRNA addition (Fig. 1c). While depletion of MPP8 led to a potent IFN response, periphilin (PPHLN1)-depletion induced a significant but lesser response, and TASOR-depleted cells only exhibited a minor shift in ISRE-GFP reporter expression. Controls had no effect on the reporter (Fig. 1c and Supplementary Fig. 1a). Interestingly, combining TASOR and MPP8 depletions revealed that TASOR-depletion could partially inhibit the endogenous IFN-stimulated gene (ISG) response mediated by MPP8-depletion (Fig. 1d). However, induction of the ISRE-GFP reporter was more potent in TASOR-depleted cells compared to control cells each treated with IFN- β (Fig. 1e). This suggests that TASOR-depleted cells are poised for innate immune activation.

The HUSH complex controls type I IFN induction. We then employed MPP8-depletion to pursue the mechanism further and first verified that the ISRE-GFP reporter was activated upon MPP8-depletion using additional MPP8 shRNAs (Supplementary Fig. 1a). We confirmed that endogenous ISGs were induced upon MPP8-depletion in three additional human cell lines: HeLa cells, THP-1 cells (acute monocytic leukemia cell line) and untransformed primary foreskin fibroblasts (HFFs) (Fig. 1f and Supplementary Fig. 1b). Supernatant from MPP8-depleted 293 cells was sufficient to activate the ISRE-GFP reporter cells in an IFN bioassay (Fig. 1g) and MPP8-depletion led to an induction of IFN- β itself, shown here by employing an IFN- β reporter cell line in which GFP is driven by an IFN- β promoter³⁶ and by IFN- β ELISA (Fig. 1h, i). We found that expression of the ISRE-GFP reporter was dependent on JAK/STAT signaling by employing a JAK/STAT inhibitor (Fig. 1j and Supplementary Fig. 1c for the endogenous response). PPHLN1-depletion led to a significant induction of type I IFN and JAK/STAT-dependent ISG expression (Supplementary Fig. 1d–f), albeit to a lesser extent than MPP8-depletion. Finally, an IFN bioassay revealed that IFN is also produced by primary fibroblasts upon MPP8 or PPHLN1-depletion (Fig. 1k). These data suggest that the HUSH complex regulates the induction and response to type I IFNs.

The HUSH complex regulates genes involved in inflammation.

To understand the global phenotype of HUSH-depletion, we performed total RNA-sequencing of primary fibroblasts. Results for MPP8-depletion showed a clear upregulation of genes involved in immunity and inflammation (Fig. 2a, b, dataset 1 in the Source data file), including 204 ISGs, which typically function in innate and adaptive immunity (employing a list we defined before in ref. 15). Of the fraction that were not defined as ISGs, 16 were KZNFs in accordance with the known function of the

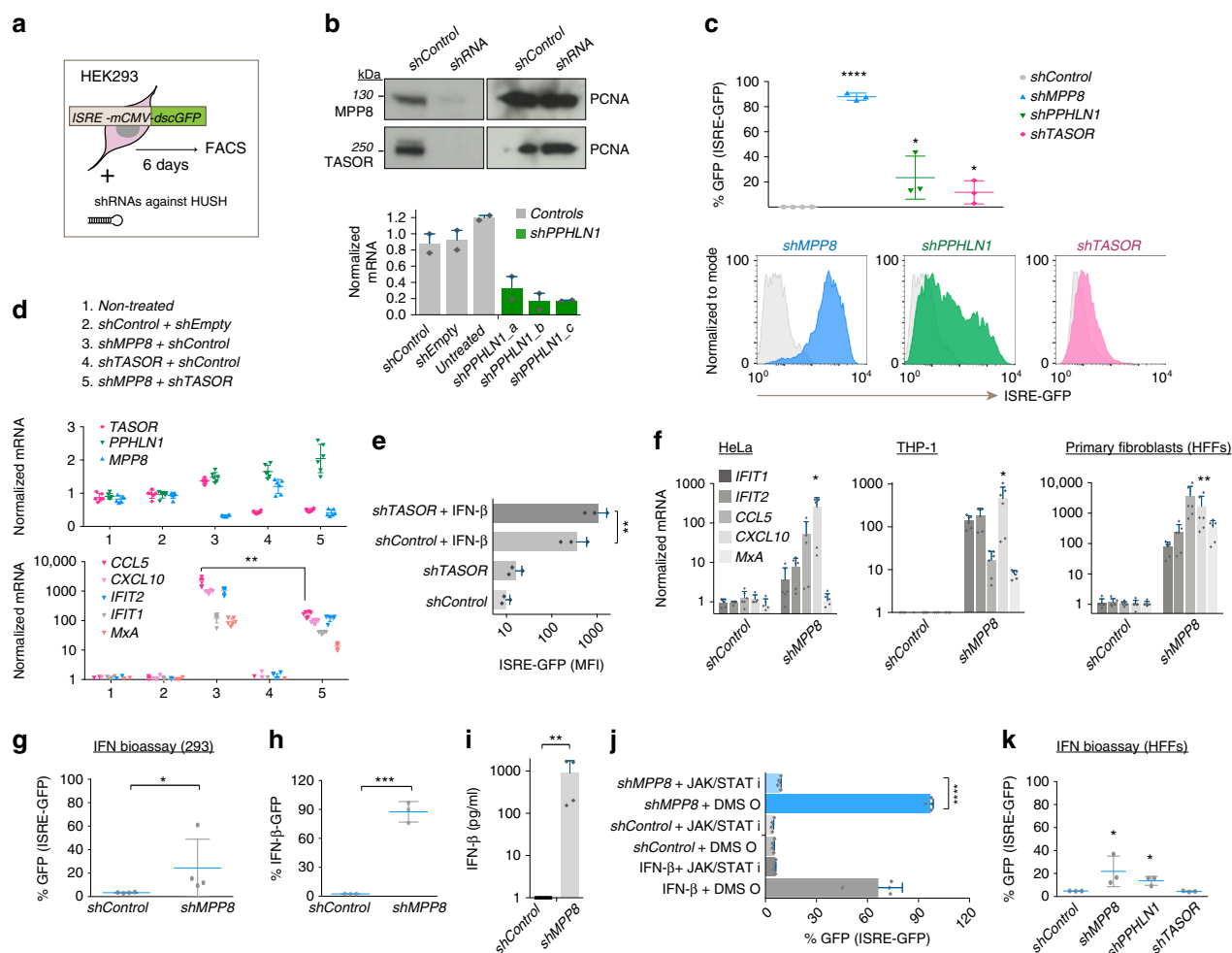


Fig. 1 The HUSH complex regulates the type I IFN response. **a** Diagram of IFN induction assay: HEK293 cells harbor an integrated IFN reporter expressing destabilized GFP (dscGFP). ISRE: IFN-stimulated response elements; *mCMV*: minimal CMV promoter. GFP was assessed 6 days post shRNAs. **b** Western blot and qRT-PCR to assess HUSH depletion day 4 post shRNAs. *PPHLN1* mRNA normalized to *GAPDH*. *N* = 3 biologically independent samples with technical qRT-PCR duplicates shown. **c** ISRE-GFP induction (6 days post shRNAs). *N* = 3 biologically independent experiments. Two-tailed unpaired *t* tests of shRNAs vs. *shControl*: $p < 0.0001$ (*shMPP8*), $p = 0.0372$ (*shPPHLN1*) and $p = 0.0466$ (*TASOR*). **d** *shMPP8* was combined with *shTASOR* or controls as stated. qRT-PCR of endogenous ISG mRNA expression (*GAPDH* normalized). *N* = 3 biologically independent experiments with technical duplicates shown. A two-tailed unpaired *t* test was used to compare *shMPP8* + *shControl* to *shMPP8* + *shTASOR* for *CCL5* ($p = 0.0021$). **e** *shTASOR* or *shControl*-treated cells were treated with IFN- β and ISRE-GFP measured 24 h later. *N* = 3 biologically independent experiments. *P* value: 0.0037 (two-tailed paired *t* test). MFI: mean fluorescent intensity. **f** QRT-PCR of ISG expression (*GAPDH* normalization) 6 days post shRNAs (or 3 days in the case of HeLa cells). *N* = 3 biologically independent experiments with technical duplicates shown. Two-tailed paired *t* tests were used to compare controls to *shMPP8* for *CXCL10*: $p = 0.0473$ (HeLa), $p = 0.0229$ (THP-1s) and $p = 0.0084$ (HFFs). **g** IFN bioassay: supernatants from shRNA-treated 293 cells were added to ISRE-GFP reporter 293 cells and GFP expression measured 24 h later. *N* = 4 biologically independent samples. $P = 0.0240$ (two-tailed paired *t* test). **h** 293 cells containing an IFN- β -destabilized GFP reporter were transduced with the stated shRNAs and GFP measured 6 days later. *N* = 3 biologically independent samples. $P = 0.0005$ (two-tailed paired *t* test). **i** Supernatants were harvested 6 days post shRNAs and used for an IFN- β ELISA. *N* = 4 biologically independent samples. $P = 0.0023$ (two-tailed paired *t* test). **j** 293 cells treated with shRNAs in the presence of the JAK/STAT inhibitor Ruxolitinib or DMSO (added from day 1) were harvested for GFP FACS at day 6. *N* = 4 biologically independent samples. IFN- β + Ruxolitinib served as a positive control. $P < 0.0001$ (two-tailed unpaired *t* test). **k** IFN bioassay: supernatants from shRNA-treated HFFs were added to ISRE-GFP reporter 293 cells and GFP expression measured 24 h later. *N* = 3 biologically independent samples. Two-tailed paired *t* test *p* values: $p = 0.0469$ (*shMPP8*), $p = 0.0320$ (*shPPHLN1*). All data presented in this figure show mean values \pm SD except (b), which depicts mean values \pm SEM.

HUSH complex in binding to and regulating KZNF genes^{25,26,33} and most of these were previously identified to be MORC2-bound³³ (Fig. 2b and Supplementary Fig. 2). Other upregulated genes were enriched for roles in chromatin regulation, cell cycle and DNA damage (Supplementary Fig. 3a–c). Of note, the IFN and inflammatory gene signature upon MPP8-depletion was reproduced in HeLa cells (Supplementary Fig. 3d). We used gene set enrichment analyses to assess the phenotype observed upon

depletion of individual HUSH components in primary fibroblasts (Fig. 2c, dataset 1 in the Source data file). *TASOR*-depleted samples differed from MPP8 and PPHLN1 in their lack of an activated IFN signature but overlapped in terms of TNF- α signaling and G2M cell cycle checkpoint. This might indicate that greater depletion of *TASOR* than that achieved here was required to observe a more complete phenotype. Concordance between PPHLN1- and MPP8-repressed genes was evidenced by a large

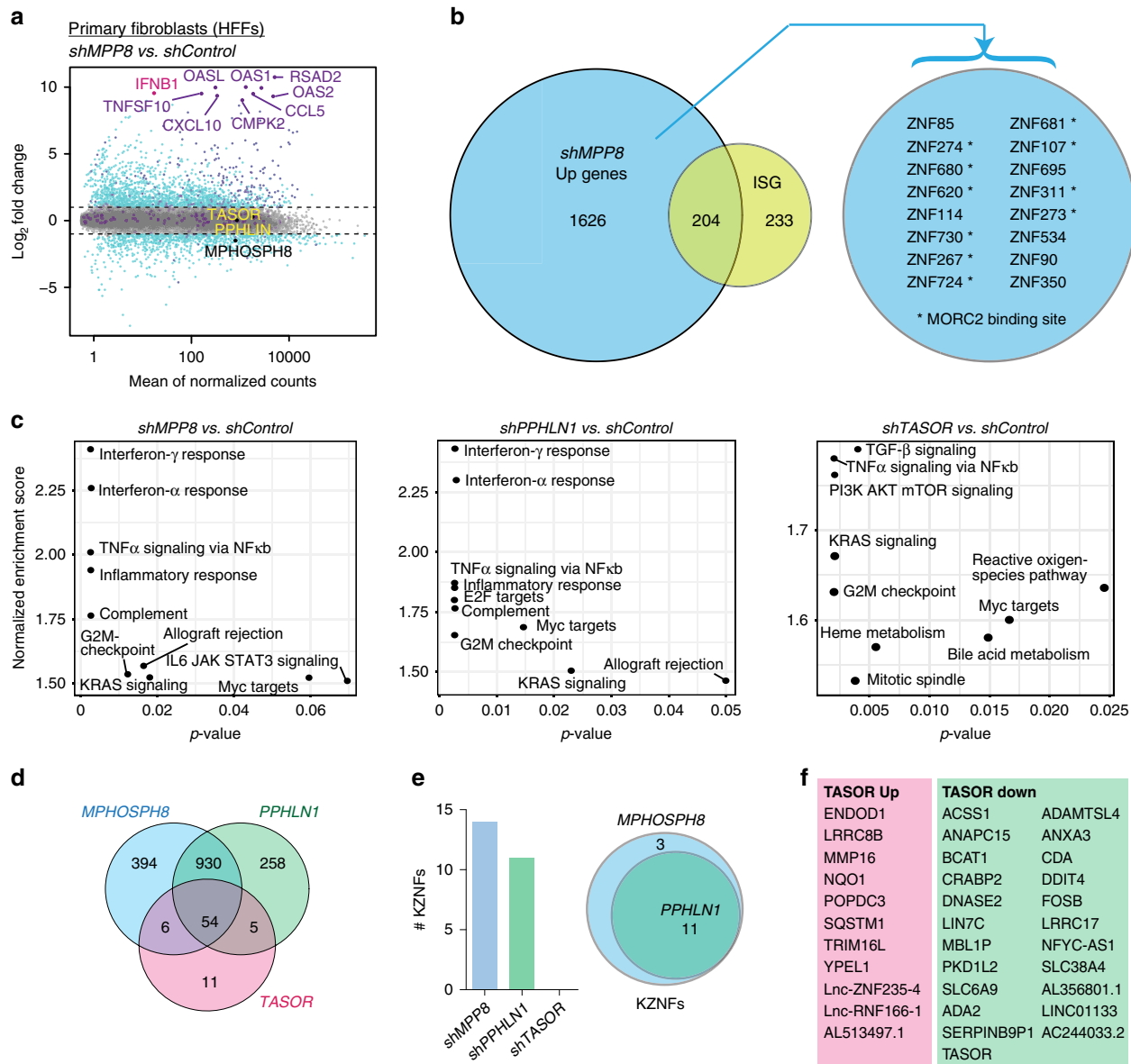


Fig. 2 The HUSH complex regulates genes involved in inflammation. Primary fibroblasts (HFFs) were treated with shRNA vectors, and RNA harvested 6 days later for total RNA-sequencing. *N* = 6 biologically independent experiments. **a** MA plot showing results of differential expression analysis, where significantly differentially expressed genes (\log_2 fold change > 1 and a *p*-adjusted value < 0.05 after Benjamini-Hochberg multiple testing correction of Wald test *p*-value of shMPP8/shControl) have been highlighted in dark cyan and ISGs are shown in purple. See dataset 1 in Source data for all exact *p* values. Example upregulated ISGs are labeled as well as HUSH components (of which MPHOSPH8/MPP8 is downregulated). **b** Venn diagram showing overlap of ISGs (list defined in ref. ¹⁵) with genes upregulated in MPP8-depleted samples. Upregulated KZNFs are listed. See also Supplementary Fig. 2 for MORC2 binding data using public ChIP data³³. **c** Total RNA-seq data were generated for samples depleted for each HUSH component vs. shControls (*n* = 3 biologically independent experiments and data are represented as gene set enrichment analyses using fgsea, where nominal *p*-values were calculated for 1000 permutations and corrected for multiple-testing using the Benjamini-Hochberg method). **d** Venn diagram showing overlap of genes upregulated in each HUSH-depletion. **e** Comparison of number and overlap of KZNFs upregulated in HUSH-depleted cells. **f** Significantly differentially-expressed genes unique to TASOR-depleted samples are listed.

number of mutually-upregulated genes (Fig. 2d), which included KZNFs for MPP8 and PPHLN1 but not TASOR (Fig. 2e). Genes uniquely up or downregulated in TASOR-depleted cells are depicted in Fig. 2f and interestingly, several of these (SQSTM1 and ACSS1) have immunomodulatory roles^{37,38}.

MPP8-depletion is accompanied by RLR-dependent RNA sensing. There is mounting evidence that epigenetic pathways prevent TEs from nucleic acid sensing through RLRs, particularly

in cancer cells^{8,10,13,14}. We detected expression of components of the RLR-signaling pathway, MDA5, RIG-I, and MAVS in all cell lines in which we had observed an IFN response upon HUSH-depletion (Fig. 3a). cGAS, in contrast, was not detected in 293 cells following IFN- β treatment, although it was readily expressed in THP-1 cells in which it was upregulated upon IFN- β treatment (Fig. 3a). We confirmed that the IFN response was MAVS-dependent using shRNA depletion of MAVS in ISRE-GFP and IFN- β GFP 293 reporter cells and we verified that MAVS-depleted cells could no longer react to stimulation with poly(I:C),

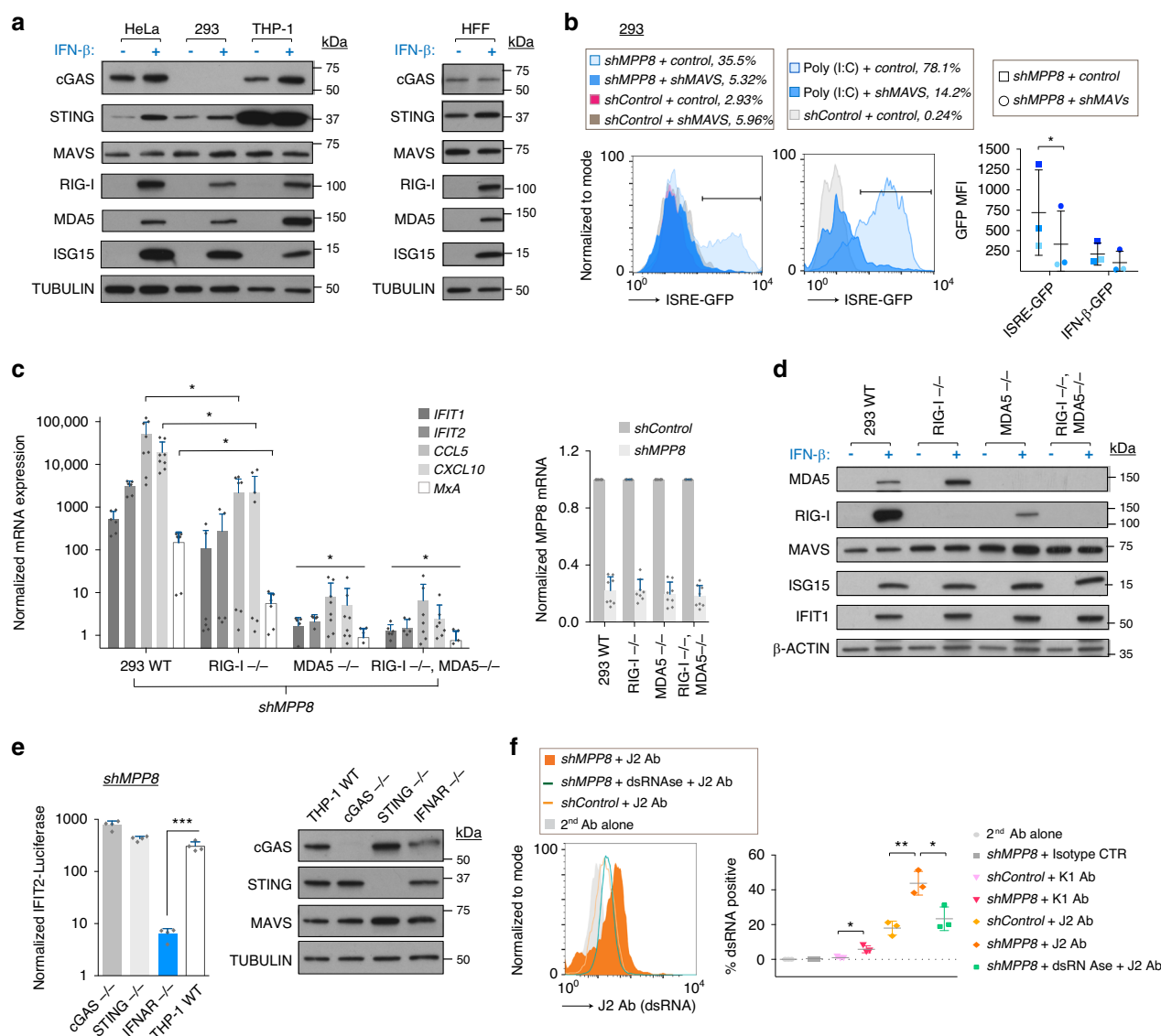


Fig. 3 MPP8-depletion is accompanied by RLR-dependent RNA sensing. **a** Western blots on extracts from stated cells treated \pm IFN- β for 24 h. $N = 2$ independent blots with one representative blot shown. **b** 293 cells were co-transduced with *shControl* or *shMPP8* plus *shMAVS* or a control vector of the same backbone (*hygromycin*) and GFP was measured on day 6 (left). MAVS depletion was verified using poly(I:C), added 24 h before FACS (middle). Summary data shown for experiments in ISRE-GFP and IFN- β -GFP reporter cells (right, $n = 3$ biologically independent samples). $P = 0.0372$ (one-tailed paired t test). MFI: mean fluorescent intensity. **c** Left: qRT-PCR expression of endogenous ISGs in the stated shRNA-treated 293 cell lines at day 6 (*GAPDH* normalized). $N = 4$ biological independent experiments with technical duplicates shown. CCL5 and CXCL10 were measured in all experiments while the other 3 ISGs were measured in 3 of the experiments. One-tailed paired t tests showed expression of all ISGs was significantly lower in MDA5-/- cells and double knockout (DKO) cells than in wildtype (WT) controls. IFIT1: $p = 0.0063$ (MDA5-/-), $p = 0.0026$ (DKO); IFIT2: 0.0007 (MDA5-/-), 0.0019 (DKO); CCL5: 0.0002 (MDA5-/-), 0.0028 (DKO); CXCL10: 0.0001 (MDA5-/-), 0.0007 (DKO); MxA: 0.0427 (MDA5-/-), 0.0423 (DKO). Expression of CCL5, CXCL10, and MxA was significantly decreased in RIG-I-/- cells compared to controls. CCL5: $p = 0.0183$; CXCL10: $p = 0.0372$; MxA: $p = 0.0139$. Right: MPP8 depletion efficiency summary ($n = 4$ biologically independent experiments with technical duplicates shown). **d** Western blots on extracts from cell lines from **c** treated \pm IFN- β for 24 h. $N = 2$ independent blots with one representative blot shown. **e** Left: stated knockout or control (WT) THP-1 reporter cell lines were treated with *shRNAs* and assayed for secreted *Luciferase* (driven by IFIT-2) 6 days later. $N = 4$ biologically independent experiments. *Luciferase* readings were normalized to background values in *shControl*-treated cells. $P = 0.0003$ (one-tailed paired t test). Right: western blots on extracts from stated cells. $N = 3$ independent blots with one representative blot shown. **f** 293 cells were treated with *shRNAs* and fixed 6 days later for intracellular staining with dsRNA antibodies, J2/K1 or control antibodies. $N = 3$ biologically independent samples. Two-tailed unpaired t test p values = 0.0166 (*shMPP8* vs. *shControl* for K1 Ab); 0.0048 (*shMPP8* vs. *shControl* for J2 Ab); 0.0212 (*shMPP8* + J2 Ab \pm dsRNAse). All data presented in this figure show mean values \pm SD.

a dsRNA analog (Fig. 3b). To dissect whether MAVS signaling was mainly dependent on RIG-I or MDA5, we employed CRISPR/Cas9-generated knockout 293 cells for either or both dsRNA sensors. Results showed that the response is dependent on

MDA5 with a contribution from RIG-I (Fig. 3c). We verified by western blot that expression of MDA5 and RIG-I were detected in the expected cells upon their treatment with IFN- β (Fig. 3d). Note that MDA5 is more highly expressed in RIG-I knockout cells than

control cells, potentially to compensate for RIG-I loss, while in MDA5 knockout cells, RIG-I expression is low, possibly contributing to the low IFN response in these cells upon MPP8-depletion (Fig. 3d). Importantly, we verified that all wildtype and knockout cell lines could mount a response to IFN- β treatment by measuring the induction of IFIT1 and ISG15 (Fig. 3d).

We then asked if the cGAS/STING pathway was necessary for IFN signaling upon MPP8-depletion in a cell line efficient for DNA sensing, (Fig. 3a). For this, we employed commercially-available THP-1 cells that stably express an IFIT2-responsive *luciferase* reporter that are either knockouts for the DNA sensors cGAS or STING, or for the IFN type-I receptor (IFNAR). The observed IFN response was independent of cGAS/STING but dependent on signaling through the type I IFN receptor, as expected (Fig. 3e). Knockout of DNA sensors was verified by western blot and functionally by stimulation with agonists (Fig. 3e right and Supplementary Fig. 3e). Since results suggested that the IFN response is dependent on dsRNA (Fig. 3b–d), we tested whether we could detect dsRNA in MPP8-depleted cells by employing antibodies (J2 and K1) that recognize dsRNA. DsRNA was apparent in MPP8-depleted cells and the signal was abrogated by pretreatment with the long dsRNA-specific endoribonuclease, RNase III (Fig. 3f).

MPP8-depletion results in overexpression of LINE-1s and LTRs. We and others previously showed that the HUSH complex mediates epigenetic repression of LINE-1 elements^{32–34}, and we hypothesized that LINE-1 RNAs could be driving the IFN response in MPP8-depleted cells. However, the HUSH complex also represses MLV²⁹ and potentially also HIV^{25,30,31}. We, therefore, first verified that our shRNA lentivector itself was not necessary to drive the ISG response, by using siRNAs against MPP8 to show that there is still induction of the ISRE-GFP reporter (Fig. 4a). This suggested that the ligand driving the response is endogenous.

We then verified that LINE-1s are overexpressed in MPP8-depleted cells by LINE-1 ORF1-protein western (Fig. 4b). We employed our RNA-sequencing data to determine which TE subfamilies are differentially expressed in MPP8-depleted cells (using TEtranscripts: TEcount tool³⁹). We found mainly primate-specific LTR (long terminal repeat) families of between 20 and 50 million years old to be derepressed, including ERV9 and HERVH (Fig. 4c and Supplementary Fig. 4a, b, dataset 2 in the Source data file). We then mapped which precise TE loci were affected in the human genome by assigning uniquely-mapping reads to the genome and intersecting results with coordinates of TEs (using RepeatMasker annotation) and we found mainly LINE-1s, SINEs, and LTRs were upregulated (Fig. 4d, e and Supplementary Fig. 4b, dataset 3 in the Source data file). Although numerous LINE-1 loci were overexpressed, they were not significantly upregulated at the subfamily level using TEtranscripts (dataset 2 in the Source data file). This suggests that LINE-1 upregulation is context-dependent, consistent with the known role of HUSH in regulating TEs in introns of genes^{33,34}. Employing available HUSH-binding data revealed that LINE-1s rather than LTRs were likely direct binding targets (Fig. 4d, e and Supplementary Fig. 4b). We highlighted the top five LINE-1 subfamilies in orange containing the most upregulated and most HUSH-bound copies, respectively (Fig. 4d, top and bottom), which spanned families between 12 and 100 million years old.

We next looked at the distance of upregulated LINE-1s and upregulated LTRs to the transcriptional start sites (TSS) of upregulated ISGs, hypothesizing that these TEs may directly regulate these genes (Fig. 4f). The LTRs were significantly closer to ISGs than to random genes suggesting that they may have been

co-opted to act as promoters or enhancers for these genes, thereby contributing to the increased IFN-responsiveness in HUSH-depleted cells (Fig. 1). The LINE-1s were not significantly close to the TSS of ISGs but were enriched within long noncoding RNAs (Supplementary Fig. 4c).

L1PA1 and L1PA2 are a potential source of dsRNA. We and others have previously shown that HUSH represses young species-specific LINE-1s, albeit mainly in early development^{32–34}. We, therefore, focused on the hominid-specific families, L1PA1 (also called L1HS) and L1PA2, which we noticed were expressed bidirectionally and are known to be MORC2-bound at the 5' UTR³³. While L1PA1 is difficult to map at the locus level, due to its high copy number and integrants being mostly identical, we mapped RNA-sequencing reads to the L1PA1 consensus sequence and found that there is significantly more production of sense and antisense strands in MPP8-depleted cells (Fig. 4g). This was also the case for L1PA2, which we show is bidirectionally-expressed as well (Supplementary Fig. 4d). Importantly, in an unbiased approach, we also identified regions in the genome with the highest (top 15%) overlapping expression in both strands and found that L1PA1 was the most overrepresented TE in those regions, followed by L1PA2. Bidirectional LINE-1s (199 L1PA1 and 74 L1PA2) were full-length with a median length of around 6 kb (Supplementary Fig. 4e) and could potentially pair in *cis* or *trans* to produce dsRNAs to accumulate in the cytoplasm. Example bidirectionally-transcribed LINE-1s and upregulated primate-conserved LINE-1s are shown (Fig. 4h, i, Supplementary Fig. 5). Of note, bidirectional L1PA1 loci were 100% identical to the ORF1 protein probed for by western blot, suggesting that we are detecting L1PA1 expression by western (Fig. 4b, dataset 4 in the Source data file).

LINE-1s regulate the type I IFN response. Data indicated that the striking IFN response observed upon MPP8-depletion could result from a mixture of components, including LINE-1 RNAs, effects of ZNF transcription factors and unveiled LTRs (Figs. 2, 4 and Supplementary Fig. 6a). We focused on the role of LINE-1 RNAs: We first determined that reverse transcription was not necessary to drive IFN by using reverse transcriptase inhibitors that are effective against LINE-1 RNA and HIV, used here as a positive control (Supplementary Fig. 6b). In contrast, overexpression of ZNF91 and to a lesser extent ZNF93, which are transcriptional repressors of SVA and LINE-1 (L1PA4) elements, respectively⁴⁰ and their co-repressor, KAP1 could partially reduce the IFN response (Supplementary Fig. 6c). The modest effect of ZNF93 might relate to the fact that it does not target young LINE-1s, which may be the main source of dsRNA (Fig. 4g and Supplementary Fig. 4d). We, therefore, designed a series of LINE-1 shRNAs and tested them to ascertain if any could inhibit the HUSH-dependent IFN response. We found that two hairpins could partially block the ISRE-GFP reporter response, one designed on the 5'UTR of L1PA1 using the consensus sequence, and one designed on L1PA4 ORF2 using an MPP8-bound element³³ (see Fig. 5a). Additionally, we designed a hairpin against MOV10, an RNA helicase thought to sequester LINE-1 RNA away from MDA5^{24,41}. This hairpin led to an augmentation of the IFN response, further suggesting that endogenous LINE-1 RNAs can mediate MDA5-sensing (Fig. 5a). We verified that the LINE-1 shRNAs could also block the endogenous ISG response (Fig. 5b). The shL1PA4 hairpin induced a detectable decrease in global expression of L1PA2-4 RNA by qRT-PCR (Fig. 5c). The L1PA1-2 hairpin recognized L1PA1s and L1PA2s equally (584 LINE-1 integrants, of which 288 were L1PA1 and 271 were L1PA2). Using primers that could detect 76 L1PA2s targeted by



Fig. 4 MPP8-depletion results in overexpression of LINE-1s and LTRs. **a** ISRE-GFP reporter 293 cells were transfected with control (CTR) or MPP8 siRNAs or dsRNA as a positive control and GFP measured on day 6. $N = 5$ biologically independent samples or $N = 3$ for the positive control. $P = 0.0108$ (two-tailed paired t test). Data show mean \pm SD. MFI: mean fluorescent intensity. **b** Left: HeLa cells were harvested 4 days post shRNAs and used for western blot for L1 ORF1 protein and alpha-Tubulin. Right: western quantification. $N = 3$ biologically independent samples. $P = 0.0426$ (one-tailed paired t test). Data show mean \pm SD. **c** Bubble plots showing the relative prevalence of all TEs in the genome by clade (orange) and relative abundance of overexpressed TE families in *shMPP8* samples compared to controls (blue) using RNA-seq data from Fig. 2 and the software Tetranscripts. See Supplementary Fig. 4a for complete diagram. Monkey images drawn by lab members. **d** Upper panel: TE loci up or downregulated in MPP8-depleted cells (log2 fold change >1 , and p adjusted values <0.05 after Benjamini-Hochberg multiple testing correction of Wald test p -value of *shMPP8/shControl* from DESeq2). See dataset 3 in Source data for all exact p values. Results are shown for loci within the stated LINE-1 families. Ns = expression not significantly changed. Lower panel: strip plot showing HUSH-bound (orange) vs. not known to be bound (gray) LINE-1 loci with their relative expression in *shMPP8* samples compared to controls. HUSH peaks defined using epic2, see Supplementary Fig. 3a, using data from ref. ³³. The approximate ages of these LINE-1 families are given underneath using ⁷⁹. See Supplementary Fig. 4b for LTR data. The subfamily names with the highest number of upregulated loci (top) or bound loci (bottom) are highlighted in orange. Boxes represent 1st and 3rd quartile, where the central line corresponds to the median; whiskers are $\times 1.5$ of the interquartile range. **e** Pie charts showing upregulated TE loci (left) and HUSH-bound TE loci (right) classified by family. **f** Median distance of Upregulated L1s ($n = 1494$) or LTRs ($n = 1177$) to transcription start sites of upregulated ISGs ($n = 204$) compared to the same number of random genes or to the median distance across 1000 randomizations. The median distance for LTRs is significantly closer than randomized genes ($p = 0.001$, where a one-sided p value was calculated as $(1 + \text{no. of randomizations with a median greater than the observed value})/1000$ randomizations). Boxplots as in (**d**) where notches indicate the confidence interval. **g** Left: the LIPA1 consensus sequence (repBase) was mapped with RNA-seq reads in both strands (left y axis) or with MORC2 ChIP and total input reads from public data ³³ (right y axis). Right: The number of RNA-seq reads mapping forward or reverse strands of the LIPA1 consensus sequence were plotted for all samples. Two-tailed unpaired t test values: $p = 0.0006$ (forward) and $p = 0.0291$ (reverse). Data show mean \pm SD. **h** Bigwig visualization of example HUSH-bound young LINE-1 expressed bidirectionally and a HUSH-sensitive primate-conserved LINE-1 expressed in sense (**i**). RNA-seq track scales were all set at 50 and normalized tracks of HUSH-binding were made using data from ref. ³³.

MPP8 is frequently downregulated in cancers. Given the emerging role of epigenetic regulators as therapeutic targets in cancer ^{42,43}, we tested if MPP8 is differentially expressed in the context of this group of diseases. For this, we used the human cancer genome atlas (TCGA) ⁴⁴ data, focusing on tissues for which there is also available expression data in healthy control samples from GTEx ⁴⁵, as well as matched TCGA controls. MPP8 was significantly downregulated in 10 out of 15 diverse human cancers as compared to both controls (Fig. 6a). We did not see the same trend for TASOR or PPHLN1 (Supplementary Fig. 6e), which may relate to the more striking phenotype observed of MPP8 depletion (Figs. 1 and 2). To gain insight into whether MPP8 depletion in cancers is associated with an IFN signature, we divided cancers by recently-defined immune subtype categories ¹⁷, plotting MPP8 expression levels. We found that MPP8 was most significantly downregulated in cancers with an IFN- γ dominant C2 immune subtype (Fig. 6b). This included high expression of the ISGs, CCL5 and CXCL10.

Given the observed trend across most cancer types to downregulate MPP8, we considered that one of the selective advantages this could confer could be a decreased capacity to repair DNA, also hinted at by the secondary genes upregulated upon MPP8 depletion, which are enriched for DNA damage, cell cycle and chromatin maintenance pathways (Supplementary Fig. 3a–c). This, together with the known risk of active LINE-1 elements in inducing DNA damage and compromising DNA repair ^{21,46–48}, led us to examine if depleting MPP8 in 293 cells was sufficient to lead to DNA damage. We observed that at baseline, MPP8-depleted cells display more damage compared to control cells, as measured by γ H2AX staining, and are also compromised in DNA repair following etoposide-induced damage (Fig. 6c). We observed the same trend in HeLa cells (Supplementary Fig. 6f).

Discussion

Here we reveal that the HUSH complex controls type I IFN signaling in human cells and that the mechanism involves dsRNA sensing by MDA5 and RIG-I. We link this cell-intrinsic innate immune response partly to LINE-1 elements, which become derepressed at the RNA and protein level following HUSH-

depletion. Surprisingly, we find that full-length hominid-specific LINE-1 elements, including L1PA1 and L1PA2 produce bidirectional transcripts. This suggests that young LINE-1s are a natural source of dsRNAs of 6Kb in length, ideal candidates for recognition by the long dsRNA sensor, MDA5, which detects dsRNA greater than 2 kb in length ⁴⁹. These RNAs were transcribed in control cells and increased in abundance in MPP8-depleted cells. We speculate that these LINE-1 dsRNAs may be chromatin-tethered in control cells and only become cytoplasmic and available for MDA5-sensing upon HUSH-depletion. Of note, we have focused on LINE1s in this study but there are likely many other potential sources of dsRNAs.

Of interest, we identify an increase in LINE-1 mRNAs derived from ancient primate-conserved LINE-1 loci in MPP8-depleted cells. These conserved LINE-1s may have evolved beneficial roles in host gene regulation since many are positioned within introns of genes as noted before ^{33,34}. Interestingly, these conserved LINE-1s are also enriched in long non-coding RNAs, which can derive from LINE-1s ⁵⁰. LINE-1 elements originated more than 160 million years ago ⁵¹ and have co-evolved with mammals to regulate genes: LINE-1 and LINE-2 elements have been repurposed in the formation of chimeric transcripts driven by LINE-1 antisense promoters ⁵², in chromatin-associated non-coding RNAs ⁵³ and in exonization events, including into immunoregulatory molecules ⁵⁴.

While the potent type I IFN response observed upon MPP8-depletion is likely to be multifactorial, as described for the immune response evoked upon treatment with pharmacological inhibitors of DNA methylation ⁴², here, we could implicate LINE-1 RNAs and potentially LINE-1-containing transcripts to play a causative role using LINE-1 knockdown and overexpression experiments (see Fig. 7 for a proposed summary model). Importantly, we also found LTRs nearby ISGs to become derepressed upon HUSH-depletion suggesting that the HUSH complex may control type I IFN responsiveness, as well as IFN induction. Of note, ERV9 LTRs that we observed to be derepressed have been described to act as cryptic promoters under epigenetic regulation ⁵⁵. LTRs have also been shown to drive production of novel cancer-specific antigens ⁵⁶.

Finally, we note that MPP8 is frequently downregulated in cancers, which may relate to the increase in DNA damage that we

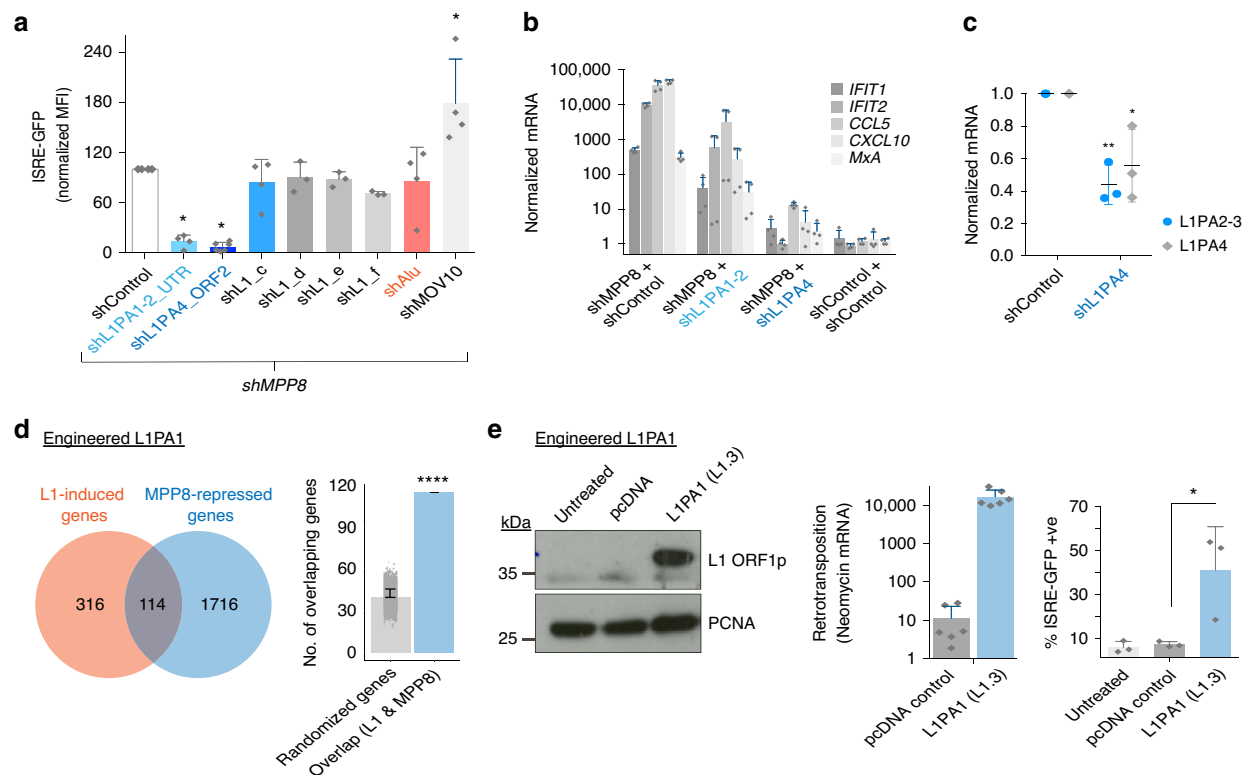


Fig. 5 LINE-1s regulate the type I IFN response. **a** 293 reporter cells were *shControl*/*shMPP8* treated together with the stated shRNAs against TE families or *shControl* again. GFP MFI (mean fluorescence intensity) at day 6 was normalized to the *shMPP8* + *shControl* sample that was set at 100%. $N = 4$ biologically independent samples, except for *shL1_d-f*, where $n = 3$ and *shL1PA4*, where $n = 6$. The control was significantly different to the *shL1PA1* group ($p = 0.0172$), the *shL1PA4* group (0.0110) and the *MOV10* group ($p = 0.0265$) (two-tailed paired t tests). **b** Groups were setup as in (**a**) but here endogenous ISG expression was measured by qRT-PCR. $N = 2$ biologically independent experiments with technical duplicates shown. **c** Knockdown efficiency was verified by qRT-PCR using global LINE-1 primers. $N = 3$ biologically independent experiments. A decrease in global LINE-1 expression was detectable for *shL1PA4*: $p = 0.0013$ (*L1PA2-3*) and $p = 0.0268$ (*L1PA4*) (two-tailed unpaired t tests). **d** Venn diagram showing overlap of genes induced by overexpression of an engineered L1PA1 reporter in RPE cells from ref. ²¹ (a Tet-inducible codon-optimised L1PA1 vs. Luciferase control that was induced for 3 days) and genes upregulated upon MPP8-depletion here. Of 114 genes that overlap, 49 are ISGs. Right: significance was assessed by selecting 10,000 randomizations of the same number (1830) of random genes and looking at the mean number that are shared with the L1-induced genes. $P = 5.1203 \times 10^{-24}$ (hypergeometric test). **e** 293 reporter cells transfected with a pcDNA control or engineered L1PA1 reporter plasmid (ksCMV-101: L1PA1 subtype L1.3). Transfection and retrotransposition verified by western blot and qRT-PCR for spliced Neomycin expression at day 4 (left and middle). $N = 3$ biologically independent experiments with technical duplicates shown. Right: IFN reporter induction was measured by GFP FACS. $N = 3$ biologically independent experiments. $P = 0.0332$ (one-tailed paired t test). All data presented in this figure show mean values \pm SD.

observe in MPP8-depleted cells. Whether DNA damage is related to IFN induction⁵⁷ or caused by DNA breaks mediated by LINE-1-encoded endonuclease⁴⁷ warrants future investigation. It will also be interesting to determine whether MPP8-depletion contributes to carcinogenesis in vivo. Our results fit with a more general emerging role for LINE-1 elements in driving IFNs and inflammation^{21–24,58–60}. Of note, although the cGAS/STING pathway was not necessary for the IFN response here, it may contribute to IFN release in vivo and could be upregulated and more active in HUSH-depleted cells. Open questions include uncovering exactly how LINE-1 elements are transcribed bidirectionally and why, and determining whether LINE-1 dsRNA is tethered to chromatin in control cells to mask if from MDA5 sensing. It's not clear if LINE-1 proteins play a role in nucleic acid sensing of LINE-1 dsRNA. Little is known too about how co-opted conserved LINE-1s and other TEs potentially control innate immune genes through gene-regulatory roles and if the HUSH complex is downregulated during physiological immune responses. It is tempting to speculate that LINE-1 elements may play a causative role in driving autoinflammatory diseases and may have been repurposed long ago to naturally prime or potentiate MDA5-mediated innate immunity.

Methods

Cell culture and reagents. HEK293 WT (referred to as 293) or CRISPR/Cas9-generated knockout 293 cells for either RIG-I, MDA5 or both and HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco, Thermo Fisher Scientific) supplemented with 100 U/ml penicillin/streptomycin (Gibco, Thermo Fisher Scientific) and 10% heat-inactivated fetal calf serum (FCS) and grown at 10% CO₂ at 37 °C. Cell lines were split 1:4 every 3 days using trypsin. 293 reporter cell lines harboring a lentivector expressing destabilized GFP under the control of IFN-stimulated responsive elements (ISRE) or destabilized GFP under control of an IFN- β promoter were a kind gift from Jan Rehwinkel^{35,36}. Note that these reporter constructs include a puromycin cassette, hence we did not use puromycin-selection of shRNAs in these cells, which was unnecessary due to their high transduction efficiency. Human primary foreskin fibroblasts (HFFs) were a kind gift from Matt Reeves and early-passage cells were split 1:3. shRNAs were puromycin-selected in HFFs. The THP-1 IFIT-2 luciferase reporter cell lines were purchased from Invivogen and cultured according to the manufacturer's guidelines. These cells are WT or CRISPR/Cas9 knockout for cGAS, STING or IFNAR and express secreted Lucia luciferase under control of a minimal IFIT2 promoter in conjunction with 5 ISRE units. IFN- β (PeproTech) or Universal type I IFN (Interferon- α/β , PBL Assay Science) were used at 10 ng/ml or 200 U/ml, respectively. Poly (I:C) (Invivogen) was used at 150 ng per well in a 24 well plate and dsRNA was made by in vitro transcription⁶¹ and used at 1 μ g/ml complexed with Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific) at a 1:3 ratio according to the manufacturer's instructions. The reverse transcriptase inhibitors, AZT and 3TC were used at 100 μ M and 200 μ M, respectively. The JAK-STAT inhibitor Ruxolitinib (Cambridge Bioscience) was used at 2 μ M, which was determined by titration on ISRE-GFP treated with IFN- β .

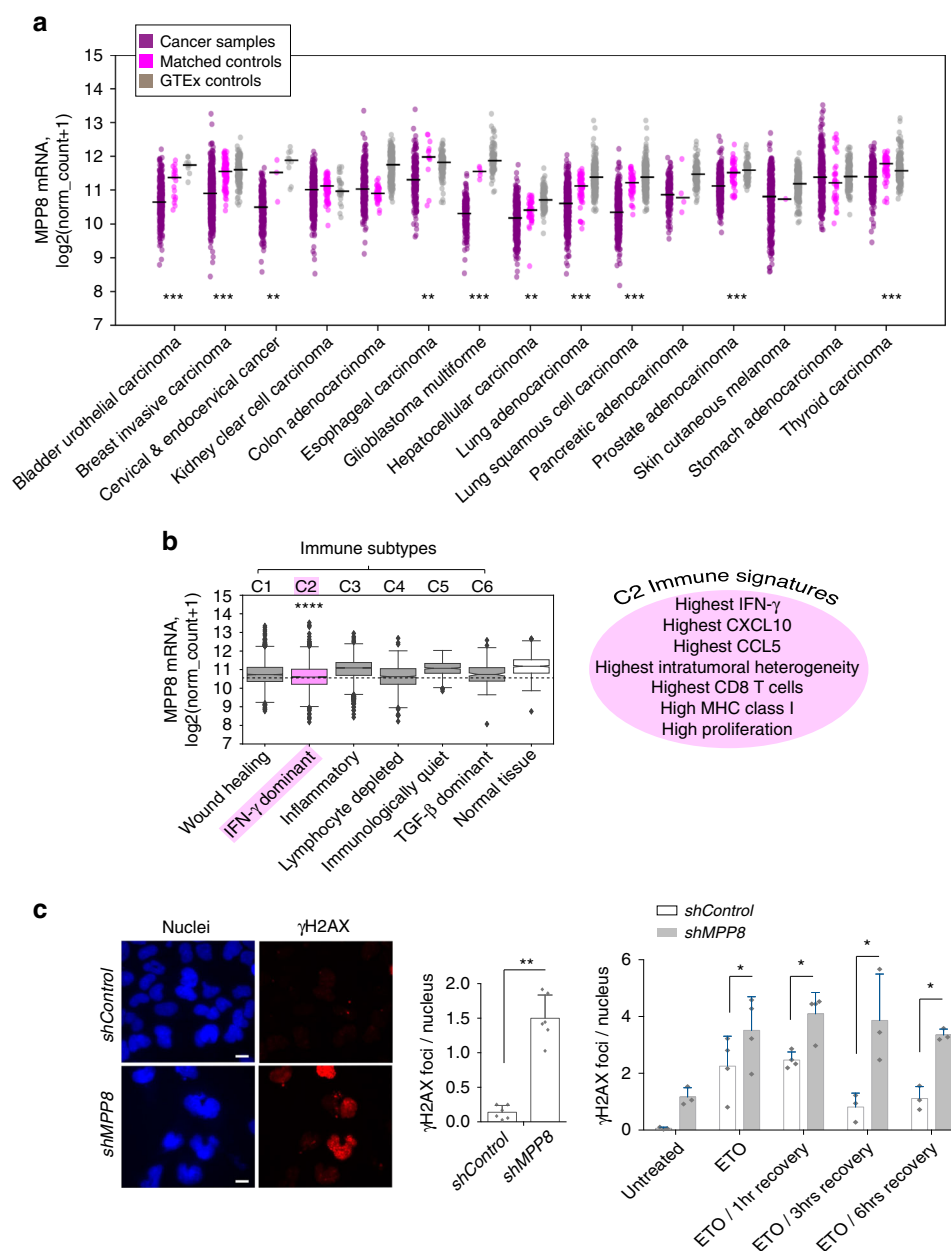


Fig. 6 MPP8 is frequently downregulated in cancers. **a** Using The Cancer Genome Atlas (TCGA) data, we selected cancers for which there were matched control samples as well as GTEx expression data from the same tissue. MPP8 is scored here as significantly downregulated when levels are lower in the cancer compared to both controls separately, using a two-sided Mann-Whitney rank test and correcting for multiple testing (p adjusted values were all <0.05 for groups marked with asterisks). Adjusted p values for matched control comparisons and GTEx comparisons respectively are given here: bladder urothelial = 2.40×10^{-5} , 1.58×10^{-6} ; breast invasive = 5.40×10^{-35} , 8.79×10^{-61} ; cervical & endocervical = 3.12×10^{-3} , 2.37×10^{-7} ; esophageal = 1.07×10^{-3} , 4.0×10^{-15} ; glioblastoma = 2.10×10^{-8} , 6.61×10^{-40} ; liver hepatocellular = 1.31×10^{-3} , 5.57×10^{-31} ; lung adeno = 1.19×10^{-11} , 9.56×10^{-82} ; lung squamous cell = 2.40×10^{-26} , 1.41×10^{-97} ; prostate adeno = 2.06×10^{-11} , 8.27×10^{-31} ; thyroid = 1.25×10^{-10} , 9.24×10^{-17} . **b** Left: all cancers were divided by immune subtype based on ref. ¹⁷ and levels of MPP8 expression were plotted. MPP8 levels were most significantly different in subtype C2 compared to normal tissue (after pairwise, two-sided t tests between each immune subtype and normal tissue with an fdr multiple test correction: $\text{padj} = 8.03 \times 10^{-94}$; C1 $n = 2,066$, C2 $n = 2,413$, C3 $n = 2,334$, C4 $n = 1,129$, C5 $n = 378$, C6 $n = 180$ samples). Boxes represent 1st and 3rd quartile and the central line is the median. Notches indicate the confidence interval; whiskers correspond to 1.5 of the interquartile range and outliers are depicted as diamonds. Right: features of cancers within immune subtype C2 are displayed (terms from ref. ¹⁷). **c** 293 reporter cells were treated with shRNAs and then either untreated or treated with etoposide (ETO) for 30 min and allowed to recover for stated times before γH2AX staining. Left: example γH2AX staining images. Middle: summary data of γH2AX foci: $N = 6$ biologically independent samples. $p = 0.0013$ (two-tailed paired t test). Right: summary data of recovery assay. $N = 4$ biologically independent samples, except for the groups, ETO and ETO/1h, where $N = 3$. Two-tailed paired t test p values = 0.0303 (ETO) 0.0486 (ETO/1h) 0.0384 (ETO/3h) and 0.0258 (ETO/6h). Data show the mean \pm SD. See Supplementary Fig. 6e for HeLa data.

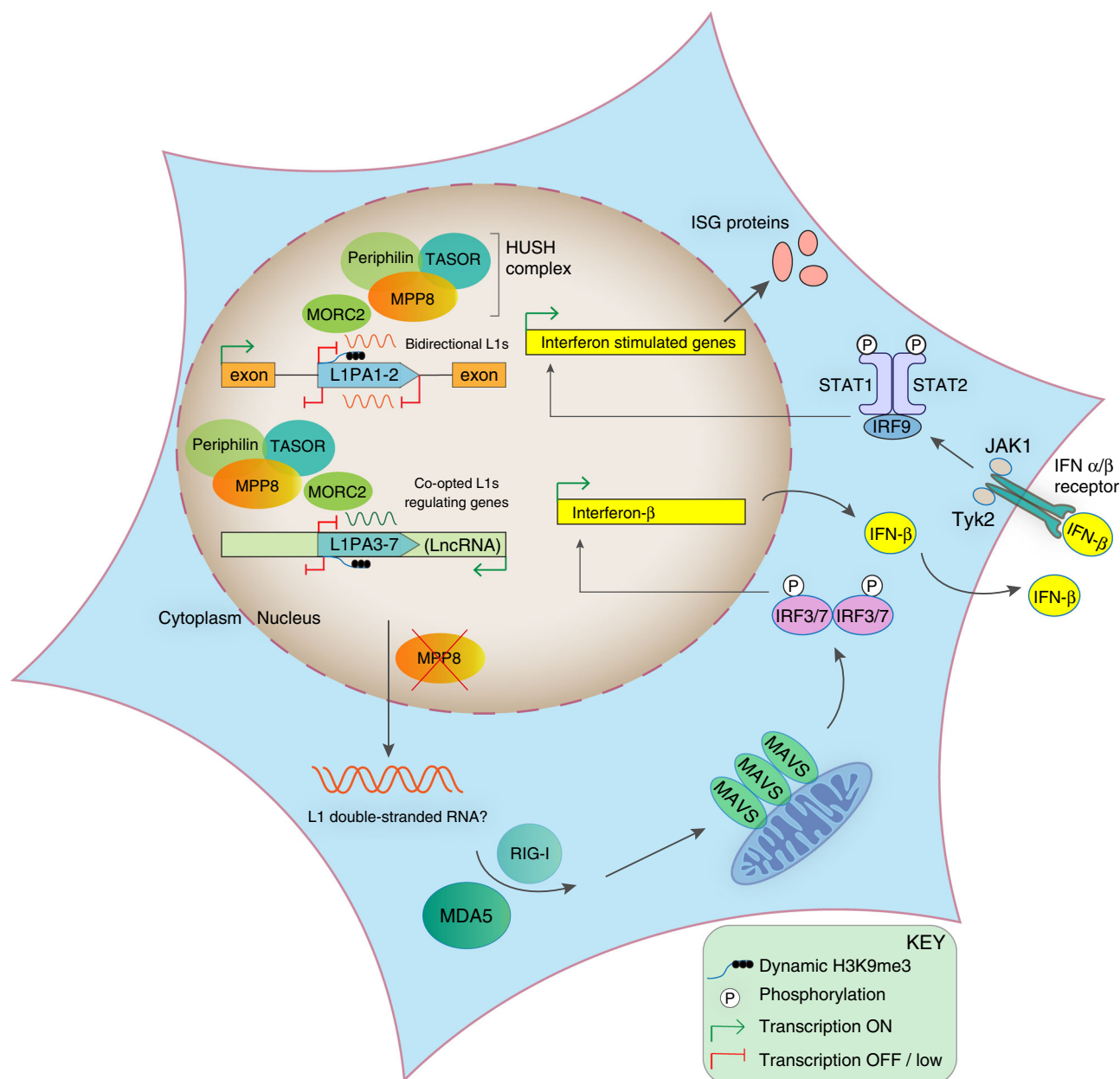


Fig. 7 Proposed model. Inactivation of the HUSH component MPP8 leads to increased expression of bidirectionally-transcribed full-length young LINE-1s (L1PA1 and L1PA2), as well as primate-conserved LINE-1s with potential co-opted roles in gene regulation, (some enriched in long non-coding RNAs). Nucleic acid sensing ensues through dsRNA sensors MDA5 and RIG-I, leading to MAVS signaling and activation and secretion of type I IFNs. IFN- α/β then act on the same cells or bystander cells to mediate induction of IFN-stimulated genes through JAK-STAT signaling. MPP8-depletion is also accompanied by DNA damage and is a factor commonly downregulated in cancers, where its decreased mRNA expression is associated with a dominant IFN immune subtype signature.

Western blotting. Cells were washed once in ice-cold PBS and lysed using pre-chilled homemade Radioimmunoprecipitation assay buffer [150 mM NaCl; 1% Triton X-100; 0.5% sodium deoxycholate; 0.1% SDS and 50 mM Tris, pH 8.0, protease inhibitors (#539134, Calbiochem) and phosphatase inhibitor cocktail (PhosSTOP, Roche)] for 30 min at 4 °C. Cell lysates were cleared by centrifugation (12,000 \times g, 20 min, 4 °C) and a fraction of lysate was used to perform a protein quantification assay (BCA Protein Assay Kit, Novagen). Standardized lysate samples were mixed with NuPAGE LDS sample buffer (Thermo Fisher) and 5% β -mercaptoethanol and heated at 95 °C for 5 min and then resolved on precast NuPAGE 4–12% Bis-Tris (Invitrogen) or handcast 10% SDS-polyacrylamide gels in tris/glycine/SDS buffer in mini-PROTEAN tanks (Biorad), followed by wet transfers onto Polyvinylidene Difluoride (PVDF) or nitrocellulose membranes, blocked in 5% non-fat dried milk in TBS-T (TBS, 0.1% Tween-20) and incubated with relevant primary antibodies: anti-MPP8 (#16796-1-AP, Proteintech), anti-TASOR/FAM208A (HPA017142, Atlas Antibodies), anti-PPHLN1 (HPA038902,

Atlas Antibodies), anti-L1 ORF1p (MABC1152, Millipore), anti-PCNA (NA03, Calbiochem), anti- α -Tubulin (T6074, Sigma Aldrich), anti-RIG-I (D14G6, #3743, Cell Signaling Technology), anti-MDA5 (D74E4, #5321, Cell Signaling Technology), anti-MAVS (#3993, Cell Signaling Technology), anti-STING (D2P2F, #13647, Cell Signaling Technology), anti-cGAS (D1D3G, #15102, Cell Signaling Technology), anti-IFIT1 (D2X9Z, #14769, Cell Signaling Technology) and anti-ISG15 (A-5, sc-166712, Santa Cruz Biotechnology). All secondary antibodies were horseradish peroxidase-conjugated (GE healthcare) and membranes were developed using ECL kits (ECL, Prime or Select kits from Amersham). Source Data are available with original blots.

shRNA lentiviral vectors and transduction. Hairpin sequences were designed against LINE-1 integrants (using the 5'UTR or ORF1 or ORF2 sequence) known to be bound by MPP8³³ (see Supplementary Table 1) and using the Clontech RNAi

designer website (<http://bioinfo.clontech.com/rnaidesigner/>) and verified in silico before cloning. Hairpin primers were annealed and then cloned into an shRNA vector¹⁵ at *Bam*HI-*Eco*RI sites and the products were checked by verification of an introduced restriction site and by sequencing. VSV-G-pseudotyped lentiviral vectors were produced by co-transfecting 293T cells in 10 cm plates with 1.5 µg of the shRNA plasmid, 1 µg p8.91 and 1 µg pMDG2 encoding VSV-G. Media was changed 1 day post transfection and supernatant harvested 48 h post-transfection and concentrated via ultracentrifugation. Puromycin selection was performed 3 days post transduction overnight or until control cells died. For 293 and HeLa cells, puromycin selection was not necessary and for this reason, dual transductions were performed in 293 cells.

CRISPR/Cas9 generation of RIG-I-like receptor knockout HEK293 cell lines.

RIG-I, MDA5, and RIG-I/MDA5 knockout 293 cells were generated by CRISPR-Cas9-mediated genome engineering. Using the CRISPR design tool provided by the Zhang lab (www.genome-engineering.org) several guide sequences were selected in human *DHX58* (encoding RIG-I) and *IFIH1* (encoding MDA5) (Supplementary Table 1). Appropriate oligonucleotides corresponding to these guide sequences were cloned into the BbsI site of pX459 (Addgene 62988), a bicistronic expression vector encoding both Cas9 and the single-guide RNA, according to the cloning protocol provided by the aforementioned website. Plasmid pX459 was obtained from the lab of Dr. Feng Zhang via Addgene (plasmid 62988)⁶². 293 cells were transfected in a 6-well plate using 1 µg pX459 per well using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Twenty-four hours post-transfection, cells were treated for 24 h with 1 µg/ml puromycin to remove untransfected cells and subsequently replated at limiting dilution in order to pick individual colonies. Knockout of RIG-I and/or MDA5 was verified by amplification and sequencing of the genomic region surrounding the CRISPR target site, immunoblotting, and in a *luciferase*-based transfection reporter assay, in which *luciferase* activity is controlled by the IFN-β promoter, using selective RIG-I and MDA5 agonists.

RNA extraction and quantification. Total RNA was extracted using RNeasy Micro kit columns (Qiagen) and DNase treated according to the manufacturer's instructions (Ambion AM1907). 500 ng of RNA was reverse transcribed using random primers and SuperScript II Reverse Transcriptase (Thermo Fisher Scientific). Control reactions were always performed in the absence of reverse transcriptase and used for qRT-PCR in parallel to cDNA to verify there was no DNA remaining. cDNA was diluted in nuclease-free water, and gene expression levels were quantified using quantitative reverse transcription PCR (qRT-PCR) using an ABI 7500 Real Time PCR System (Applied Biosystems). SYBR green Fast PCR mastermix (Life Technologies) was used. CT values for the test genes were normalized against those of *Gapdh* or *B2M* using the $-\Delta\Delta C_t$ method to calculate fold change. See Supplementary Table 2 for primer sequences, which includes a breakdown of TE primer hits to TE subfamilies.

Flow cytometry. Cells were trypsinized and harvested in media and centrifuged and then fixed in 1% paraformaldehyde for 10 min (or unfixed) and washed in PBS, resuspended in PBS and run on a BD FACSCalibur or LSRFortessa Flow Cytometer acquiring 10 000 cells per sample to meet statistical robustness using BD CellQuest or FACSDiva software. Intracellular FACS³⁴ was performed in 293 cells (not reporter cells) using the dsRNA-specific antibodies, J2 and K1 (from English and Scientific Consulting Hungary, SCICONS) or an isotype control antibody (mouse IgG2a, Santa cruz sc-3878), with a secondary Alexa Fluor 488 goat anti-mouse IgG (ThermoFisher). We verified the antibodies to recognize dsRNA by intracellular FACS after transfection of dsRNA. ShortCut RNase III was used for long dsRNA digestion (NEB). Data were analyzed using FlowJo (Tree Star version 10.3.0). The flow cytometry gating strategy, which involved gating on live cells and then on GFP positive cells (or dsRNA-positive cells) using the negative control sample is shown in Supplementary Fig. 7.

Luciferase assays. 293T cells were plated at 10^5 cells/ml in 12 well plates, 1 ml per well and co-transfected with 500 ng KZNFs, 50 ng SVA or L1PA4 *luciferase* reporter plasmids (kind gift from David Haussler⁴⁰) and 5 ng pRT-TK *Renilla* control plasmid. Cells were then lysed and *luciferase* measured 48 h later using the Dual Luciferase assay kit (Promega, E1910) and a Glomax 96 microplate Luminometer (Promega) using the Dual Glow Programme according to the manufacturer's guidelines.

Interferon-β ELISA. An ELISA kit precoated with an antibody against human IFN-β (VeriKine Human IFN Beta ELISA Kit, 41410 by PBL) was used according to the manufacturers' instructions. Absorbance measurements were taken at 450 nm using a Multiskan FC Microplate Photometer and SkanIT Software. Optical densities were plotted using a 4-parameter fit for the standard curve and this was used to determine the IFN-β concentrations of test samples.

Engineered LINE-1 reporter constructs. RNA-sequencing data were employed from ref. ²¹, in which a Tet-inducible codon-optimised L1PA1 (ORFeus) was

induced in RPE cells using doxycycline treatment for 3 days. vs. an induced *luciferase* control. A list of genes that were significantly upregulated (>10 fold with *p* adjusted values <0.05) was obtained (for data access, see: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119999>) and the overlap of this list compared to genes upregulated in this study (HFF cells) upon depletion of MPP8. For experiments in which we overexpressed engineered LINE-1, we used a full-length active L1PA1 construct expressed from a CMV promoter (ksCMV-101/L1.3, which is not codon-optimized, kind gift from Jose Garcia-Perez). This construct contains a neomycin resistance cassette, inserted within the 3'UTR, in the opposite direction, interrupted by an intron. Expression of the neomycin gene depends on transcription, splicing, reverse transcription and integration. Engineered L1PA1 was transfecting into 293 ISRE-GFP reporter cells vs. pcDNA control and 4 days later, L1 ORF1 protein, spliced Neomycin expression and ISRE-GFP reporter induction were measured.

Total RNA-sequencing and analysis. The RNA was quality checked on a 4200 TapeStation using the RNA ScreenTape assay (Agilent Technologies, Wokingham, UK) and the RNA concentration measured using a Qubit RNA Broad Range kit (Life Technologies, Paisley, UK). Total RNA samples were then processed using KAPA's stranded RNA HyperPrep RiboErase kit using an input of 500 ng per sample. Samples were sequenced on a NextSeq 500 instrument (Illumina Cambridge, Chesterford, UK) after pooling libraries in equimolar quantities, using a 2x 151 bp paired-end run, resulting in over 15 million reads per sample. Illumina's bcl2fastq Conversion Software was used to demultiplex data and generate fastq files. Fastq files were checked for quality with FastQC v.0.11.8⁶³ and TrimGalore v0.4.1⁶⁴ was used to remove adaptors only. Reads trimmed for adaptors were aligned to the human genome USCS build hg38 using Tophat 2.1.0⁶⁵ and Gencode v30⁶⁶ gene annotations (https://www.encodegenes.org/human/release_30.html). The number of read counts per gene was calculated using HTSeq-Count⁶⁷, the Bioconductor package DESeq2⁶⁸ was used to estimate differential expression of genes across conditions, and the Approximate Posterior Estimation method⁶⁹ was used to shrink the logarithmic fold change. As per the default behavior of DESeq2, *p*-values were adjusted for multiple testing with the Benjamini-Hochberg false-discovery-rate (FDR) procedure. Genes were considered as significantly differentially expressed when the adjusted *p*-values were <0.05, and where the log2 of the fold change was >1 for upregulated or < -1 for downregulated genes. BiomaRt⁷⁰ was used to convert gene identifiers and gene set enrichment analysis, based on rank derived from the $abs(\log_2FC) * (-\log_{10}(pvalue))$ and the Hallmark Gene Set Collection from MSigDB^{71,72} (<https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H>), was performed with the fgsea package 1.8.0⁷³ and gage 2.32.0 from R Bioconductor. We defined the list of ISGs before¹⁵ by employing <http://www.interferome.org> and selecting genes upregulated by 10-fold upon IFN treatment. Enrichment of Reactome pathways (<https://reactome.org/>) within gene subsets was performed with clusterProfiler⁷⁴. For analysis of individual TE integrants, TopHat2 alignments were used to obtain the number of unique reads mapping to individual TE loci using HTSeq-count and custom-made annotation files where each TE locus was given a unique identifier. To reduce the number of features provided as input for the analysis, TE loci with less than 20 counts across all 21 samples were filtered out before submitting the files with the counts per TE loci to DESeq2. Differential expression analysis was performed as for genes and including gene counts. Venn diagrams were made with VennDiagram 1.6.20.

For analysis of TE family expression, the TEcounts tool of Tetranscripts³⁹ was used to obtain TE counts, followed by differential expression analysis using DESeq2. TE families were considered to be significantly differentially expressed when the adjusted *p*-values were <0.05, and where the log2 of the fold change was >1 for upregulated families and < -1 for downregulated families. HUSH binding to TEs was assessed using public data³³ found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95374>.

Expression coverage tracks were generated with the *genomecov* tool from bedtools⁷⁵, scaled by library size, where read pairs were separated by their strand of origin and individual replicates were merged for knockdown and control samples of biological triplicates from the same batch. Different vector preparations (produced at different times) were used for each biological replicate. bigWig tracks were visualized using the Integrative Genomics Viewer⁷⁶. For identification of bidirectional transcription, genome-wide coverage was calculated for 500 bp bins, using the above-mentioned bigWig tracks. A bin was considered to be expressed from a given strand if its mean expression was above the value equivalent to the 85th percentile of all bins with an expression value greater than 0. Bins that passed the threshold for both strands were considered to be bidirectionally transcribed. Bidirectionally transcribed bins were merged into larger intervals when found adjacent to each other using bedtools merge. We then intersected results with RepeatMasker annotation (<http://www.repeatmasker.org/>) to uncover bidirectional TEs, with the top hit being L1PA1.

For the TEs with the largest number of loci within bidirectionally-transcribed bins (L1PA1 and L1PA2), we retrieved consensus sequences from RepBase (<https://www.girinst.org/repbase/>) and aligned all reads to each individual reference (full-length L1PA1, full-length L1PA2, assembled from RepBase fragments: L1P1_5end, L1P1_orf2, L1PA2_3end, or the sequence of a full-length L1PA2 previously identified to be MPP8-bound³³, coordinates: chr1: 207733982-207734498). Depth across the different sequences was calculated using samtools depth, normalized

using size factors calculated for the entire library. Total number of reads mapping to a reference was obtained with the samtools flagstat command.

Each TE family in the human genome was assigned a taxonomy level using information from Dfam (<https://dfam.org/>). The relative abundance per taxonomy level of all families in the genome or upregulated families only was visualized in a phylogenetic tree obtained from the NCBI Taxonomy Database (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>) using the ETE Toolkit⁷⁷.

The distance between the TE subsets (L1s or LTRs upregulated in shMPP8 samples compared to shControl samples) and the transcription start sites (TSSs) of IFN response genes was calculated using the pybedtools closest function with RepeatMasker annotation and TSSs retrieved using biomaRt (<https://www.ensembl.org/biomart/martview/>). Random subsets of genes were generated within python. Total RNA-sequencing data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135765>.

The cancer genome atlas (TCGA) and GTEx analysis. Phenotype data were downloaded from <https://xenabrowser.net/> for the TCGA TARGET GTEx cohort⁷⁸. The phenotype table was manually curated to include only cancer types that had both a matched normal tissue from TCGA and an independent normal tissue from GTEx. The *TcgaTargetGtex_RSEM_Hugo_norm_count* dataset was queried using the xenaPython python API. Differences in expression between cancer and control samples were statistically assessed using a Mann-Whitney *U* test followed by an FDR multiple test correction in python. The results shown in Fig. 6 are in part based upon data generated by the TCGA research network (<https://www.cancer.gov/tcga>)⁴⁴, The Genotype-Tissue Expression (GTEx) Project⁴⁵ and The TARGET Program (<https://ocg.cancer.gov/programs/target>)⁷⁸. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS; the data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.vN.pN on 01/09/2019. Raw data from Synapse ICGC-TCGA Whole Genome Pan-Cancer Analysis project (<https://www.synapse.org/#!Synapse:syn2351328/wiki/62351>, <https://www.synapse.org/#!Synapse:syn11288411>) was downloaded and visualised using UCSC Xena (<https://doi.org/10.1101/326470>)⁷⁸. Immune subtype data were retrieved from https://xenabrowser.net/datapages/?dataset=Subtype_Immune_Model_Based.txt&host=https%3A%2F%2Fpancanatlas.xenahubs.net&addHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu&removeHub=https%3A%2F%2Fpcawg.xenahubs.net. Source data are available with RNA-seq datasets.

DNA damage. HEK293 or HeLa cells were transfected with shRNAs and then 2×10^4 cells plated per well two days later in 96 well Cell-carrier optical plates. 4–6 days post transduction, cells were treated with 60 μ M or 30 μ M etoposide for 20 min or left untreated. For recovery studies, etoposide-treated cells were washed 3x in PBS and new media was added for the indicated times. Cells were fixed with 3% paraformaldehyde, quenched with 50 mM NH₄Cl and permeabilised with 0.3% Triton X-100 in PBS. Blocking was performed in PBS containing 1% FCS before washing and staining with the γ H2AX primary antibody for 1 hour and then washed and stained for 1 h with a Red Alexa Fluor-conjugated secondary antibody. Cells were then washed and stained with DAPI. Images were acquired and analyzed using a Hermes WiScan-automated system (IDEA Bio-Medical Ltd. Rehovot, Israel), using ImageJ software.

Statistical analysis. All data are presented with error bars showing standard deviation (SD) or standard error of the mean (SEM), where stated and statistical significance was assessed using two tailed, unpaired Student *t* tests, or other statistical tests where stated (see figure legends for details) employing GraphPad prism. The number of biological replicates is stated in the figure legends and technical replicates where relevant: for flow cytometry technical replicates were 10,000 events. A *P*-value of <0.05 was considered statistically significant (*****p* < 0.0001, ****p* < 0.001, ***p* < 0.01 and **p* < 0.05).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Total RNA-sequencing data are available on the NCBI Gene Expression Omnibus database (accession number, [GSE135765](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135765)) and other data are included in this article and its supplementary information files. Accession numbers for the publicly-available data are: [GSE95374](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95374) for ChIP-sequencing data on the HUSH complex, and [GSE119999](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119999) for RNA-sequencing data on cells expressing engineered LINE-1s. All data is available from the authors upon reasonable request. Source data are provided with this paper.

Code availability

Code is deposited in the following open source repository (https://github.com/regmdr/HUSH_analysis) and detailed in the methods and all code can be retrieved by contacting the corresponding author.

Received: 13 August 2019; Accepted: 28 September 2020;

Published online: 03 November 2020

References

- Platt, R. N. II., Vandeweghe, M. W. & Ray, D. A. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* **26**, 25–43 (2018).
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384 (2011).
- Frank, J. A. & Feschotte, C. Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* **25**, 81–89 (2017).
- Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
- Castro-Diaz, N. et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
- Garcia-Perez, J. L. et al. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466**, 769–773 (2010).
- Jonsson, M. E. et al. Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* **10**, 3182 (2019).
- Chiappinelli, K. B. et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
- Collins, P. L., Kyle, K. E., Egawa, T., Shinkai, Y. & Oltz, E. M. The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc. Natl Acad. Sci. USA* **112**, 8367–8372 (2015).
- Cuellar, T. L. et al. Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *J. Cell Biol.* **216**, 3535–3549 (2017).
- Ecco, G. et al. Transposable elements and Their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev. Cell* **36**, 611–623 (2016).
- Kato, M., Takemoto, K. & Shinkai, Y. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. *Nat. Commun.* **9**, 1683 (2018).
- Roulois, D. et al. DNA-Demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* **162**, 961–973 (2015).
- Sheng, W. et al. LSD1 Ablation stimulates anti-tumor immunity and enables checkpoint blockade. *Cell* **174**, 549–563 e19 (2018).
- Tie, C. H. et al. KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. *EMBO Rep* **19**, e45000 (2018).
- Streich, F. & Jouvenet, N. Stimulation of innate immunity by host and viral RNAs. *Trends Immunol.* **40**, 1134–1148 (2019).
- Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830 e14 (2018).
- Ahmad, S. et al. Breaching self-tolerance to Alu duplex RNA underlies MDA5-mediated inflammation. *Cell* **172**, 797–810 e13 (2018).
- Chung, H. et al. Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **172**, 811–824 e14 (2018).
- Ishizuka, J. J. et al. Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **565**, 43–48 (2019).
- Ardeljan, D. et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat. Struct. Mol. Biol.* **27**, 168–178 (2020).
- De Cecco, M. et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).
- Thomas, C. A. et al. Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. *Cell Stem Cell* **21**, 319–331 e8 (2017).
- Zhao, K. et al. LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. *J. Autoimmun.* **90**, 105–115 (2018).
- Tchakovnikarova, I. A. et al. Gene silencing. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348**, 1481–1485 (2015).
- Tchakovnikarova, I. A. et al. Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat. Genet.* **49**, 1035–1044 (2017).
- Marnef, A. et al. A cohesin/HUSH- and LINC-dependent pathway controls ribosomal DNA double-strand break repair. *Genes Dev.* **33**, 1175–1190 (2019).
- Douse, C. H. et al. Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nat. Commun.* **9**, 651 (2018).

29. Zhu, Y., Wang, G. Z., Cingoz, O. & Goff, S. P. NP220 mediates silencing of unintegrated retroviral DNA. *Nature* **564**, 278–282 (2018).
30. Chougui, G. et al. HIV-2/SIV viral protein X counteracts HUSH repressor complex. *Nat. Microbiol.* **3**, 891–897 (2018).
31. Yurkovetskiy, L. et al. Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. *Nat. Microbiol.* **3**, 1354–1361 (2018).
32. Fukuda, K., Okuda, A., Yusa, K. & Shinkai, Y. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res.* **28**, 846–858 (2018).
33. Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232 (2018).
34. Robbez-Masson, L. et al. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* **28**, 836–845 (2018).
35. Bridgeman, A. et al. Viruses transfer the antiviral second messenger cGAMP between cells. *Science* **349**, 1228–1232 (2015).
36. Hertzog, J. et al. Infection with a Brazilian isolate of Zika virus generates RIG-I stimulatory RNA and the viral NS5 protein blocks type I IFN induction and signaling. *Eur. J. Immunol.* **48**, 1120–1136 (2018).
37. Qiu, J. et al. Acetate promotes T cell effector function during glucose restriction. *Cell Rep.* **27**, 2063–2074 e5 (2019).
38. Xian, Y., Yang, S., Jin, S., Zhang, Y. & Cui, J. LRRC59 modulates type I interferon signaling by restraining the SQSTM1/p62-mediated autophagic degradation of pattern recognition receptor DDX58/RIG-I. *Autophagy* **16**, 408–418 (2020).
39. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
40. Jacobs, F. M. et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242–245 (2014).
41. Warkocki, Z. et al. Uridylation by TUT4/7 restricts retrotransposition of human LINE-1s. *Cell* **174**, 1537–1548 e29 (2018).
42. Jones, P. A., Ohtani, H., Chakravarthy, A. & De Carvalho, D. D. Epigenetic therapy in immune-oncology. *Nat. Rev. Cancer* **19**, 151–161 (2019).
43. Tie, C. H. & Rowe, H. M. Epigenetic control of retrotransposons in adult tissues: implications for immune regulation. *Curr. Opin. Virol.* **25**, 28–33 (2017).
44. Cancer Genome Atlas Research, N. et al. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
45. Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
46. Belgnaoui, S. M., Gosden, R. G., Semmes, O. J. & Haoudi, A. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.* **6**, 13 (2006).
47. Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **357**, 1383–1393 (2006).
48. Mita, P. et al. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat. Struct. Mol. Biol.* **27**, 179–191 (2020).
49. Kato, H. et al. Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-1 and melanoma differentiation-associated gene 5. *J. Exp. Med.* **205**, 1601–1610 (2008).
50. Zhao, T. et al. LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol.* **19**, 195 (2018).
51. Ivancevic, A. M., Kortschak, R. D., Bertozzi, T. & Adelson, D. L. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.* **19**, 85 (2018).
52. Miglio, U. et al. The expression of LINE1-MET chimeric transcript identifies a subgroup of aggressive breast cancers. *Int. J. Cancer* **143**, 2838–2848 (2018).
53. Percharde, M. et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391–405 e19 (2018).
54. Ng, K. W. et al. Soluble PD-L1 generated by endogenous retroelement exaptation is a receptor antagonist. *Elife* **8**, e50256 (2019).
55. Brocks, D. et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* **49**, 1052–1060 (2017).
56. Attig, J. et al. LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. *Genome Res.* **29**, 1578–1590 (2019).
57. Gao, L. et al. Bone marrow-derived mesenchymal stem cells from patients with systemic lupus erythematosus have a senescence-associated secretory phenotype mediated by a mitochondrial antiviral signaling protein-interferon-beta feedback loop. *Arthritis Rheumatol.* **69**, 1623–1635 (2017).
58. Bregnard, C. et al. Upregulated LINE-1 activity in the Fanconi anemia cancer susceptibility syndrome leads to spontaneous pro-inflammatory cytokine production. *EBioMedicine* **8**, 184–194 (2016).
59. Jung, H., Choi, J. K. & Lee, E. A. Immune signatures correlate with L1 retrotransposition in gastrointestinal cancers. *Genome Res.* **28**, 1136–1146 (2018).
60. Simon, M. et al. LINE1 Derepression in aged wild-type and SIRT6-deficient mice drives inflammation. *Cell Metab.* **29**, 871–885 e5 (2019).
61. Maillard, P. V. et al. Inactivation of the type I interferon pathway reveals long double-stranded RNA-mediated RNA interference in mammalian cells. *EMBO J.* **35**, 2505–2518 (2016).
62. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
63. Anders, S. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010)
64. Krueger, F. Trim Galore. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
65. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
66. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
67. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
69. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
70. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
71. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
72. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
73. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*. Preprint at <https://doi.org/10.1101/060012> (2019).
74. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
75. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
76. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
77. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
78. Goldman, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).
79. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).

Acknowledgements

We thank Jose Garcia-Perez for the engineered L1PA1 construct, Jan Rehwinkel for the ISRE & IFN- β reporter HEK293 cell lines, Matthew Reeves for the primary fibroblasts and Paul Lehner and Caetano Reis e Sousa for advice. We thank the Pathogen Genomics Unit at UCL run by Judy Breuer for total RNA-sequencing runs, and the TCGA and GTEx consortiums for public expression data^{44,45}. This work was funded through a Sir Henry Dale Fellowship through the Wellcome Trust and Royal Society (Grant number 101200/Z/13/Z, supporting H.M.R., H.T., C.H.C.T., and J.H.) and a European Research Council starting grant (678350, TransposonsReprogram, supporting R.E.-G., P.G., L.F., and H.M.R.). P.M. and R.G. are funded by a Wellcome Trust Senior Fellowship to R.G. (WT108082AIA). P.V.M. received funding through a Wolfson UCL Excellence Fellowship and is funded through a UKRI Future Leaders Fellowship (MR/S034498/1). This work was also supported through a consumables grant funded by the Rosetrees Trust and Robert Luff Foundation (A2630) and a Barts Charity Lectureship (MMBG1R) awarded to H.M.R. This work is dedicated to the memory of Paul Rowe.

Author contributions

H.T., R.E.-G., C.H.C.T., and H.M.R. conceived, designed, and performed experiments, analyzed data and wrote the paper. P.G., P.M., R.G., L.F., J.H., A.G.V., E.G., K.H.B., and P.V.M. performed experiments and contributed to ideas. All authors read and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19170-5>.

Correspondence and requests for materials should be addressed to H.M.R.

Peer review information *Nature Communications* thanks Todd Macfarlan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020